

P14387-A

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11238053 A**

(43) Date of publication of application: 31 . 08 . 99

(51) Int. Cl

**G06F 17/21**

**G06F 17/24**

(21) Application number: 10041644

(22) Date of filing: 24 . 02 . 98

(71) Applicant: **MATSUSHITA ELECTRIC IND CO LTD**

(72) Inventor: **IKUTA JUNZO  
NUMATA YASUYUKI**

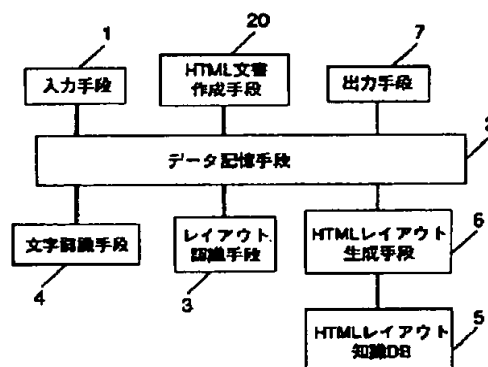
(54) **DOCUMENT DEVICE, DOCUMENT  
GENERATION METHOD AND RECORDING  
MEDIUM RECORDING DOCUMENT  
GENERATION PROGRAM**

(57) Abstract:

PROBLEM TO BE SOLVED: To convert an existed document inputted as picture data into a document described by prescribed structured tag language.

SOLUTION: Picture data inputted from an input means 1 is layout-recognized by a layout recognition means 3 and an HTML layout generation means 6 generates the control code of HTML based on respective recognized layouts and HTML(hyper text mark up language) layout knowledge 5. A character recognition means 4 character-recognizes input picture data and outputs a character code. An HTML document generation means 20 generates the HTML document by associating the control code of HTML with the character code and outputs it from an output means 7.

COPYRIGHT: (C)1999,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-238053

(43) 公開日 平成11年(1999) 8月31日

(51) IntCl.<sup>8</sup>

G 0 6 F 17/21  
17/24

識別記号

F I

G 0 6 F 15/20

5 3 0 A

5 3 4 P

審査請求 未請求 請求項の数 5 O L (全 10 頁)

(21) 出願番号 特願平10-41644

(22) 出願日 平成10年(1998) 2月24日

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 生田 淳三

大阪府門真市大字門真1006番地 松下電器  
産業株式会社内

(72) 発明者 沼田 泰之

大阪府門真市大字門真1006番地 松下電器  
産業株式会社内

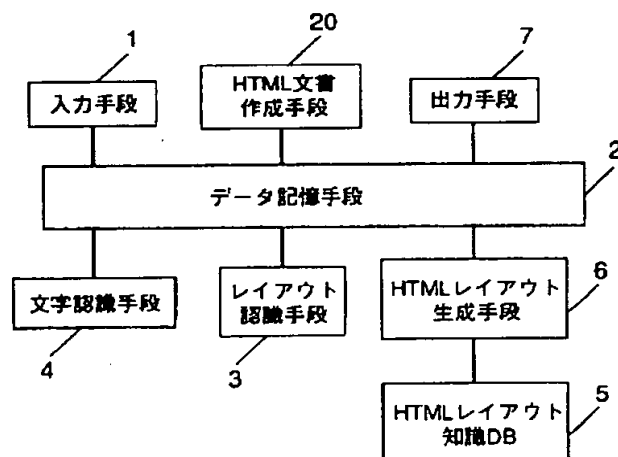
(74) 代理人 弁理士 滝本 智之 (外 1 名)

(54) 【発明の名称】 文書作成装置および文書作成方法ならびに文書作成プログラムを記録した記録媒体

(57) 【要約】

【課題】 画像データとして入力される既存の文書を、所定の構造化タグ言語で記述された文書に変換すること。

【解決手段】 入力手段 1 から入力される画像データを、レイアウト認識手段 3 によりレイアウト認識し、認識された各レイアウトと HTML レイアウト知識 DB 5 とに基づいて、HTML レイアウト生成手段 6 により HTML の制御コードを生成する。また、文字認識手段 4 は入力画像データを文字認識して文字コードを出力し、HTML 文書作成手段 20 は、HTML の制御コードと文字コードとを関連付けて HTML 文書を作成し、出力手段 7 より出力する。



## 【特許請求の範囲】

【請求項1】画像データとして入力される文書から、所定の構造化タグ言語で記述された文書を作成する文書作成装置であって、  
文書を画像データとして入力する入力手段と、  
前記入力手段から入力される画像データを記憶する入力画像記憶手段と、  
入力画像記憶手段に記憶される画像データのレイアウトを認識するレイアウト認識手段と、  
構造化タグ言語のレイアウトに関する変換ルールを記憶したレイアウト知識記憶手段と、  
前記レイアウト知識記憶手段の変換ルールに基づいて、前記レイアウト認識手段により認識される各レイアウトに対して構造化タグ言語の制御コードを生成する制御コード生成手段と、  
前記入力画像記憶手段に記憶される画像データ中の文字イメージを認識し、文字コードに変換する文字認識手段と、  
前記文字認識手段により出力される文字コードと、前記制御コード生成手段により生成される制御コードとを関連付けて、構造化タグ言語で記述された文書を作成する文書作成手段と、  
前記文書作成手段により作成された文書を出力する出力手段と、を有することを特徴とする文書作成装置。  
【請求項2】前記レイアウト知識記憶手段は、利用者により変換ルールを追加可能であることを特徴とする請求項1に記載の文書作成装置。  
【請求項3】前記文書作成手段は、複数の入力画像データから得られる各文書を連結し、1つの構造化タグ言語で記述された文書として生成することを特徴とする請求項1に記載の文書作成装置。  
【請求項4】画像データとして入力される文書から、所定の構造化タグ言語で記述された文書を作成する文書作成方法であって、  
入力される画像データを記憶するステップと、  
入力される画像データのレイアウトを認識するステップと、  
構造化タグ言語のレイアウトに関する変換ルールを参照し、認識される各レイアウトに対して構造化タグ言語の制御コードを生成するステップと、  
入力される画像データ中の文字イメージを認識し、文字コードに変換するステップと、  
変換された文字コードと、生成された制御コードとを関連付けて、構造化タグ言語で記述された文書を作成するステップと、  
作成された文書を出力させるステップと、を有することを特徴とする文書作成方法。  
【請求項5】画像データとして入力される文書から、所定の構造化タグ言語で記述された文書を作成する文書作成プログラムを記録した記録媒体であって、

入力される画像データを記憶するステップと、  
入力される画像データのレイアウトを認識するステップと、  
構造化タグ言語のレイアウトに関する変換ルールを参照し、認識される各レイアウトに対して構造化タグ言語の制御コードを生成するステップと、  
入力される画像データ中の文字イメージを認識し、文字コードに変換するステップと、  
変換された文字コードと、生成された制御コードとを関連付けて、構造化タグ言語で記述された文書を作成するステップと、  
作成された文書を出力させるステップと、を有することを特徴とする文書作成プログラムを記録した記憶媒体。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、例えばインターネット（Internet）上で、ワールド・ワイド・ウェブ（World Wide Web）のホームページを記述するための言語であるHTML（HyperText Markup Language）など、構造化タグ言語で記述された文書を作成する文書作成装置および文書作成方法ならびに文書作成プログラムを記録した記録媒体に関するものである。

## 【0002】

【従来の技術】テキストファイル中にタグと呼ばれる文字列（制御コード）を置くことで、文字の配置、大きさ、使用フォントなどを指定することのできる構造化タグ言語が知られている。構造化タグ言語で記述された文書は、ブラウザと呼ばれる専用のプログラムを用いることにより、文書に記述されている制御コードに従って表示される。

【0003】この構造化タグ言語の代表的なものとして、インターネット上のWWWのホームページを記述する言語であるHTMLが知られている。特に、近年のインターネット技術やイントラネット技術の発達ならびに普及により、様々な文書をHTML文書に変換する要求が高まっている。当然ながら、従来より保持している紙に印字された膨大な文書をHTML文書の形式に変換して利用したいという要求も大きい。

【0004】紙に印字された文書をHTML文書のような構造化タグ言語で記述された文書に変換する場合に、次の2つの方法がある。1つ目の方法は、文書を光学式スキャナ装置で読み取り、それをそのままイメージデータとしてHTML文書内に組み込んで利用する方法である。2つ目の方法は、文書を見ながら人手でHTML用の制御コードを付与していく方法である。

## 【0005】

【発明が解決しようとする課題】しかしながら、従来の方法では次のような問題点がある。まず、イメージデータとして利用する場合には、イメージデータが巨大になり、記憶する容量やユーザが閲覧するためにデータを取

得する時間が巨大になるという問題点がある。また、人手でHTML用の制御コードを付与する場合には、膨大な作業時間を必要とし、効率的ではない。

【0006】本発明は、画像データとして入力される既存の文書を、所定の構造化タグ言語で記述された文書に変換することのできる文書作成装置および文書作成方法ならびに文書作成プログラムを記録した記録媒体を提供することを目的とする。

【0007】

【課題を解決するための手段】この課題を解決するために本発明は、画像データとして入力される文書から、所定の構造化タグ言語で記述された文書を作成する文書作成装置であって、文書を画像データとして入力する入力手段と、入力手段から入力される画像データを記憶する入力画像記憶手段と、入力画像記憶手段に記憶される画像データのレイアウトを認識するレイアウト認識手段と、構造化タグ言語のレイアウトに関する変換ルールを記憶したレイアウト知識記憶手段と、レイアウト知識記憶手段の変換ルールに基づいて、レイアウト認識手段により認識される各レイアウトに対して構造化タグ言語の制御コードを生成する制御コード生成手段と、入力画像記憶手段に記憶される画像データ中の文字イメージを認識し、文字コードに変換する文字認識手段と、文字認識手段により出力される文字コードと、制御コード生成手段により生成される制御コードとを関連付けて、構造化タグ言語で記述された文書を作成する文書作成手段と、文書作成手段により作成された文書を出力する出力手段と、を有する構成とした。

【0008】これにより、画像データとして入力される既存の文書を、所定の構造化タグ言語で記述された文書に変換することのできる文書作成装置が得られる。

【0009】

【発明の実施の形態】本発明の請求項1に記載の発明は、画像データとして入力される文書から、所定の構造化タグ言語で記述された文書を作成する文書作成装置であって、文書を画像データとして入力する入力手段と、入力手段から入力される画像データを記憶する入力画像記憶手段と、入力画像記憶手段に記憶される画像データのレイアウトを認識するレイアウト認識手段と、構造化タグ言語のレイアウトに関する変換ルールを記憶したレイアウト知識記憶手段と、レイアウト知識記憶手段の変換ルールに基づいて、レイアウト認識手段により認識される各レイアウトに対して構造化タグ言語の制御コードを生成する制御コード生成手段と、入力画像記憶手段に記憶される画像データ中の文字イメージを認識し、文字コードに変換する文字認識手段と、文字認識手段により出力される文字コードと、制御コード生成手段により生成される制御コードとを関連付けて、構造化タグ言語で記述された文書を作成する文書作成手段と、文書作成手段により作成された文書を出力する出力手段と、を有す

る構成としたことにより、画像データとして入力される文書のレイアウトに従って、所定の構造化タグ言語の制御コードを付与し、構造化タグ言語で記述された文書に変換することができる。

【0010】本発明の請求項2に記載の発明は、請求項1に記載の発明において、レイアウト知識記憶手段は、利用者により変換ルールを追加可能である構成としたことにより、利用者の持つ文書に特有の表現に対応できる。

10 【0011】本発明の請求項3に記載の発明は、請求項1に記載の発明において、文書作成手段は、複数の入力画像データから得られる各文書を連結し、構造化タグ言語で記述された1つの文書として生成する構成としたことにより、物理的に複数のページにわたる文書であっても、構造化タグ言語で記述された1つのページとして利用できる。

20 【0012】以下、本発明の実施の形態について、図面を参照しながら説明する。なお、以下の各実施の形態では、構造化タグ言語の具体例として、HTMLを例にとり説明を行っている。

【0013】（実施の形態1）図1は本発明の第1の実施の形態における文書作成装置の機能ブロック図であり、機能手段による構成を示している。

30 【0014】図1において、1は画像イメージを入力する入力手段、2は画像イメージやレイアウト認識、文字認識結果などを記憶するデータ記憶手段、3は入力された画像イメージのレイアウトを認識するレイアウト認識手段、4は文字イメージを文字コードに変換する文字認識手段と、5はHTMLレイアウトに関する変換ルールを記憶したHTMLレイアウト知識DB（データベース）、6はHTMLレイアウト知識DB5にしたがって、HTMLの制御コードを生成するHTMLレイアウト生成手段、7は作成されたHTML文書をファイル等に保存したりディスプレイに表示したりする出力手段、20は文字認識手段4より出力される文字コードと、HTMLレイアウト生成手段6により生成された制御コードとを関連付けてHTML文書を作成するHTML文書作成手段である。

40 【0015】図2は本発明の第1の実施の形態における文書作成装置の装置ブロック図であり、ハードウェアによる構成を示している。

【0016】図2において、8はデータを読み込むスキャナである。9は制御プログラムを実行することにより装置を制御する中央処理演算装置（CPU）である。10はデータを記憶するリード・オンリー・メモリ（ROM）であり、CPU9が実行するプログラムを格納する領域を有する。11はデータを一時的に記憶するランダム・アクセス・メモリ（RAM）である。12は陰極線管ディスプレイ（CRT）などの表示装置である。

50 【0017】13はフロッピー・ディスク（FD）など

## 5

の記録媒体にデータを読み書きするディスク・ドライブである。14はキーボードやポインティング・デバイスなどの入力装置、15はハード・ディスクなどの2次記憶装置である。16はデータ・バスである。

【0018】ここで、図1の機能手段と図2のハードウェアとの対応関係を説明する。図1および図2に示すように、入力手段1はスキャナ8により実現される。データ記憶手段2はRAM11により実現される。HTMLレイアウト知識DB5はROM10あるいは2次記憶装置15により実現される。出力手段7はCRT12あるいはディスク・ドライブ13により実現される。

【0019】レイアウト認識手段3、文字認識手段4、HTMLレイアウト生成手段6は、CPU9がROM10およびRAM11とデータのやりとりを行いながらROM10に記憶されている制御プログラムを実行することにより実現される。

【0020】なお、本実施の形態では、CPU9はROM10に格納された制御プログラムを実行する形態を示しているが、記録媒体17からディスク・ドライブ13を介して制御プログラムを読み込み、RAM11あるいは2次記憶装置15に展開したものをCPU9が実行する形態としても良い。このような形態とすることにより、本発明を汎用コンピュータでも容易に実現可能とすることができる。

【0021】以上のように構成された文書作成装置について、以下にその動作について説明する。なお、以下のフローチャートは、CPU9がROM10に格納されている制御プログラムを実行する様子について示したものである。

【0022】図3は本発明の第1の実施の形態における文書作成装置の動作フローチャートである。

【0023】図3に示すように、まず、入力手段1から画像が入力され、入力される画像データをデータ記憶手段2に記憶させる(ステップ1)。ここでは、図5に示す画像が入力されたものとする。

【0024】次に、レイアウト認識手段3は、入力された画像イメージの解析を行い、レイアウトの認識を行う(ステップ2)。ここでは、文字領域が大きく2つに分れ、「我が家のレシピ集」がセンタリングされ、比較的大きな文字で書かれていること、および「昨日のメニュー」から続く3行が箇条書きされていることが認識される。なお、ここで「我が家のレシピ集」という書き方を行ったが、まだ文字認識は行われておらず、現時点では「我が家のレシピ集」という文字列が生成されているわけではなく、そこにあるイメージの部分のことを示している。

【0025】HTMLレイアウト生成手段6は、HTMLレイアウトの生成を行う(ステップ3)。このステップ3は、図4を使って詳しく説明する。

【0026】図4は本発明の第1の実施の形態における

## 6

レイアウト生成の動作フローチャートである。

【0027】図4に示すように、まず、ステップ301では、HTMLレイアウト生成手段6は、HTMLレイアウト知識DB5より、1つのルールを取り出す。図6にHTMLレイアウト知識DB5の例を示す。ここでは、「センタリングされている大きな文字→<CENTER><H1>と、</H1></CENTER>で囲む」が取り出される。

【0028】次に、ステップ302では、HTMLレイアウト生成手段6は、レイアウト認識手段3が出力したレイアウト情報を参照し、取り出したルールに合致するレイアウトがあるかどうか調べる。ここでは、「我が家のレシピ集」の部分がルールに一致するのでステップ303へ進む。

【0029】ステップ303では、ルール中の「→」以降の処理がなされ、「我が家のレシピ集」の部分が、<CENTER><H1>と、</H1></CENTER>で囲むことがデータ記憶手段2に記憶される。

【0030】ステップ302に戻り、再度、ルールに合致するレイアウトがあるかどうか調べる。ここでは存在しないので、ステップ304に進む。

【0031】ステップ304では、HTMLレイアウト生成手段6は、HTMLレイアウト知識DB5に記憶されているルールのうち、全てのルールの処理を行ったかどうか調べる。図6に示すルールのうち、まだ1つしか処理を終えていないのでステップ301に戻る。

【0032】ステップ301では、HTMLレイアウト生成手段6は、HTMLレイアウト知識DB5より、別のルールを取り出す。図6にHTMLレイアウト知識DB5の例を示す。ここでは、「箇条書きになっている部分→全体を<UL>と</UL>で囲み、それぞれの項目を<LI>と</LI>で囲む」が取り出される。

【0033】ステップ302では、HTMLレイアウト生成手段6は、レイアウト認識手段3が出力したレイアウト情報を参照し、ルールに合致するレイアウトがあるかどうか調べる。ここでは、「昨日のメニュー」から「明日のメニュー」までが箇条書き部分であるので、ルールに一致し、ステップ303へ進む。

【0034】ステップ303では、ルールの「→」以降の処理がなされ、「昨日のメニュー」から「明日のメニュー」までの全体を<UL>と</UL>で囲み、また、「昨日のメニュー」、「今日のメニュー」、「明日のメニュー」をそれぞれ<LI>と</LI>で囲むことがデータ記憶手段2に記憶される。

【0035】ステップ302に戻り、再度、ルールに合致するレイアウトがあるかどうか調べる。ここでは存在しないので、ステップ304に進む。

【0036】ステップ304では、HTMLレイアウト生成手段6は、HTMLレイアウト知識DB5に記憶されているルールのうち、全てのルールの処理を行ったかどうか調べる。ここでは全てのルールの処理が終わった

10

20

30

40

50

ものとして、ステップ305へ進む。

【0037】ステップ305では、処理されていないレイアウトが存在するか調べる。ここでは全てのルールが試されたが、処理されていないレイアウトが存在した場合は、ステップ306に示すごとく、レイアウト処理ができない部分があったことを出力手段7に出力する。

【0038】なお、ステップ306で警告を表示する代わりに、何らかのデフォルト的な処理を行ってもよい。例えば、全ての文字を本文のデータとして何のレイアウト情報なしに利用する、レイアウトが確認できなかったデータはコメントとして挿入する、などである。

【0039】以上で、図4に示すHTMLレイアウト生成が終了するので、図3に戻りステップ4に進む。

【0040】図3に戻り、入力画像データの文字認識を行う(ステップ4)。ステップ4では、各文字領域の文字を認識し、文字コードを出力する。文字コードが出力されると、HTML文書作成手段20は、ステップ3で生成したHTMLレイアウト情報に従って文字コードを埋め込み、HTML文書を作成する(ステップ5)。ステップ5では、さらに、HTML文書であることの必須の制御コードとして、文書の先頭に<HTML><BODY>を、また文書の最後に</BODY></HTML>をそれぞれ付与する。

【0041】作成されたHTML文書は、出力手段7により出力される。全てのデータが出力手段7に出力されている状態を図7に示す。この出力は、例えば、ディスク・ドライブ13に保存したり、表示装置12に表示したりすることである。

【0042】以上のように本実施の形態によれば、入力される画像データのレイアウト認識を行い、レイアウト認識した結果に基づいて、HTMLの制御コードを生成し、さらに、文字認識した結果と前記制御コードを関連付けることで、HTML文書を作成するので、操作者は、既存の文書をHTML文書に変換する際、制御コードを手動で付与する必要がなくなり、作業の効率化を図ることができる。

【0043】(実施の形態2)以下、本発明の第2の実施の形態について説明する。

【0044】本実施の形態では、第1の実施の形態に示すHTMLレイアウト知識DB5に利用者が必要に応じてルールを追加・修正することができるようにするものである。

【0045】例えば、図8に示すように、罫線で囲まれた3段組の表を含んだ文書をHTML文書化する場合、HTMLレイアウト知識DB5に、罫線で書かれた3段組の表を処理するルールが存在しないとする。

【0046】このような場合、操作者は、キーボードなどの入力装置を用い、図9に示すように、HTMLレイアウト知識DB5に新たなルールを追加することにより、第1の実施の形態で示した手順により図10に示すHTML文書を得ることができる。

【0047】以上のように本実施の形態では、このようにしてHTMLレイアウト知識DB5に用意していなかったようなレイアウトを持つ文書に対しても、操作者がルールを追加・修正することによりHTML文書を自動的に作成することができる。

【0048】(実施の形態3)以下、本発明の第3の実施の形態について説明する。

【0049】本実施の形態では、第1の実施の形態で示した文書作成装置において、入力される文書の複数のページを、1つのHTML文書として処理できるようにするものである。

【0050】図11は本発明の第3の実施の形態における文書作成装置の動作フローチャートである。

【0051】図11に示すように、入力手段1より複数の画像データが入力され、データ記憶手段2に記憶される(ステップ501)。

【0052】入力された画像データは、第1の実施の形態に示したステップ2～ステップ5の手順、すなわち、図3に示す手順により、HTML文書に変換される(ステップ502)。

【0053】この手順を、全てのページが処理されるまで続ける(ステップ503)。ここでは図12に示す2ページの文書をHTML文書化することを例に取り説明すると、図13に示すように、それぞれのページがHTML文書化される。

【0054】次に、HTML文書作成手段20は、複数ページの結合処理を行う(ステップ504)。具体的には、それぞれのページ間の<HTML><BODY>および</BODY></HTML>を削除して、それぞれのページをつなげる。その結果を図14に示す。

【0055】作成されたHTML文書は、出力手段7より出力される(ステップ505)。この出力は、例えば、ディスク・ドライブ13に保存したり、表示装置12に表示したりすることである。

【0056】以上のように本実施の形態によれば、論理的には1つなりのページで、用紙の都合上複数ページに分割されているため、入力される画像データが複数に分かれている場合でも、1つのHTML文書として作成することが可能となる。

【0057】以上、各実施の形態ではある形態を例にとり説明を行ったが、本発明は、上記した各実施の形態のみにとどまらないことは言うまでもない。例えば、入力手段1としてスキャナ8を例にとり説明したが、スキャナ8から入力するのではなく、画像データをディスク・ドライブ13やネットワークを介して入力する場合も考えられる。

【0058】また、出力結果を記憶媒体17に出力する例を説明したが、フラッシュ・メモ리카ードや、ネットワークを介した他のデバイスへ出力することなども考えられる。

## 【0059】

【発明の効果】以上のように本発明によれば、画像データとして入力される文書から、所定の構造化タグ言語で記述された文書を作成する文書作成装置であって、文書を画像データとして入力する入力手段と、入力手段から入力される画像データを記憶する入力画像記憶手段と、入力画像記憶手段に記憶される画像データのレイアウトを認識するレイアウト認識手段と、構造化タグ言語のレイアウトに関する変換ルールを記憶したレイアウト知識記憶手段と、レイアウト知識記憶手段の変換ルールに基づいて、レイアウト認識手段により認識される各レイアウトに対して構造化タグ言語の制御コードを生成する制御コード生成手段と、入力画像記憶手段に記憶される画像データ中の文字イメージを認識し、文字コードに変換する文字認識手段と、文字認識手段により出力される文字コードと、制御コード生成手段により生成される制御コードとを関連付けて、構造化タグ言語で記述された文書を作成する文書作成手段と、文書作成手段により作成された文書を出力する出力手段と、を有する構成としたことにより、画像データとして入力される文書のレイアウトに従って、所定の構造化タグ言語の制御コードを付与し、構造化タグ言語で記述された文書に変換することができるので、変換作業における操作者の操作負担が軽減される。

【0060】また、レイアウト知識記憶手段は、操作者により変換ルールを追加可能である構成としたことにより、利用者の持つ文書に特有の表現に対応でき、以後の変換精度を向上させることができる。

【0061】また、文書作成手段は、複数の入力画像データから得られる各文書を連結し、構造化タグ言語で記述された1つのHTML文書として生成する構成としたことにより、物理的に複数のページにわたる文書であっても、構造化タグ言語で記述された1つのページとして利用でき、変換後の利用効率化を図ることができる。

## 【図面の簡単な説明】

【図1】本発明の第1の実施の形態における文書作成装置の機能ブロック図

【図2】本発明の第1の実施の形態における文書作成装置の回路ブロック図

【図3】本発明の第1の実施の形態における文書作成装置の動作フローチャート

【図4】本発明の第1の実施の形態におけるレイアウト生成の動作フローチャート

【図5】本発明の第1の実施の形態における入力画像データを示す図

【図6】本発明の第1の実施の形態におけるHTMLレイアウト知識DBを示す図

【図7】本発明の第1の実施の形態における作成されたHTML文書を示す図

【図8】本発明の第2の実施の形態における入力画像データを示す図

【図9】本発明の第2の実施の形態におけるHTMLレイアウト知識DBを示す図

【図10】本発明の第2の実施の形態における作成されたHTML文書を示す図

【図11】本発明の第3の実施の形態における文書作成装置の動作フローチャート

【図12】本発明の第3の実施の形態における入力画像データを示す図

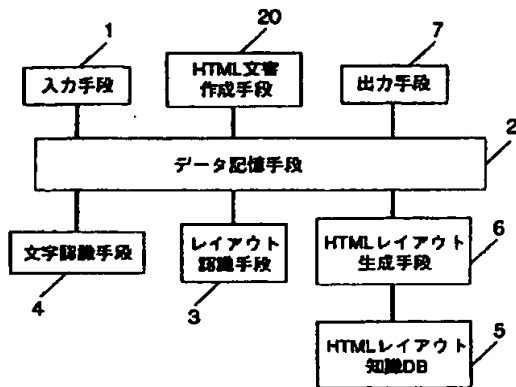
【図13】本発明の第3の実施の形態における処理途中のHTML文書を示す図

【図14】本発明の第3の実施の形態における作成されたHTML文書を示す図

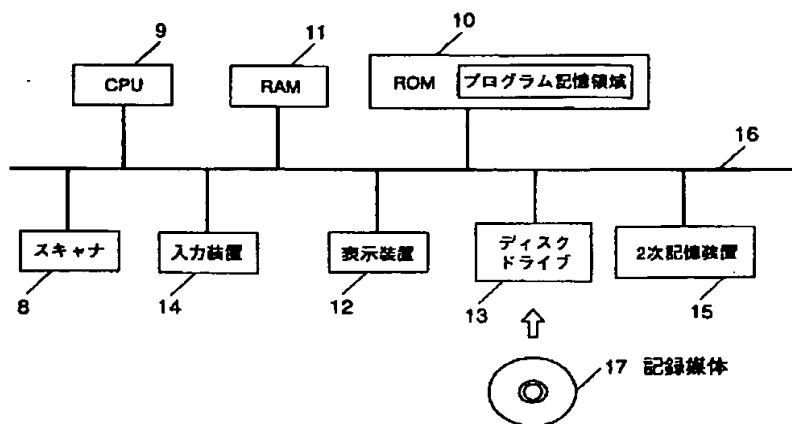
## 【符号の説明】

- 1 入力手段
- 2 データ記憶手段
- 3 レイアウト認識手段
- 4 文字認識手段
- 5 HTMLレイアウト知識DB
- 6 HTMLレイアウト生成手段
- 7 出力手段
- 8 スキャナ
- 9 中央処理装置 (CPU)
- 10 リード・オンリー・メモリ (ROM)
- 11 ランダム・アクセス・メモリ (RAM)
- 12 表示装置
- 13 ディスク・ドライブ
- 14 入力装置
- 15 2次記憶装置
- 16 データ・バス
- 20 HTML文書作成手段

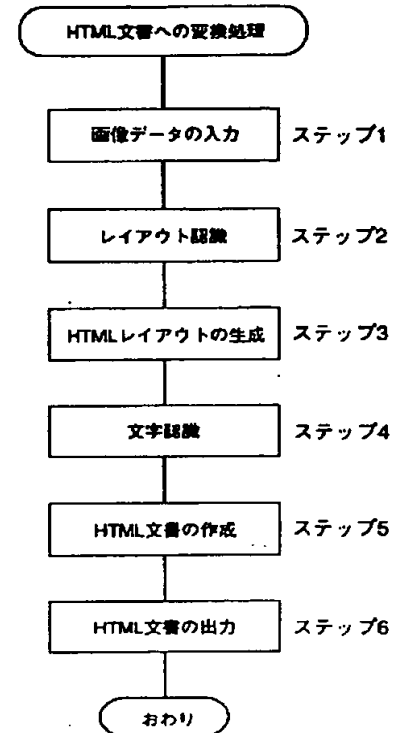
【図1】



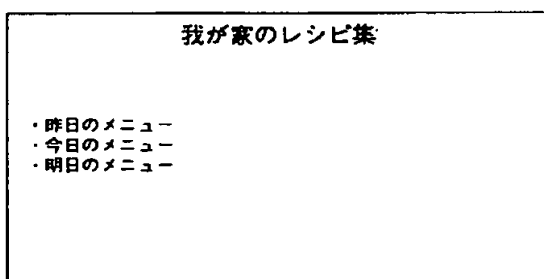
【図2】



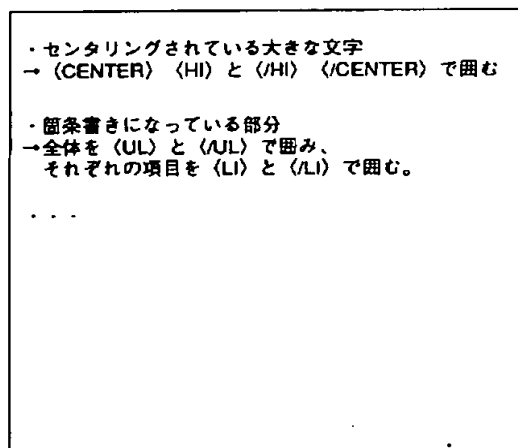
【図3】



【図5】

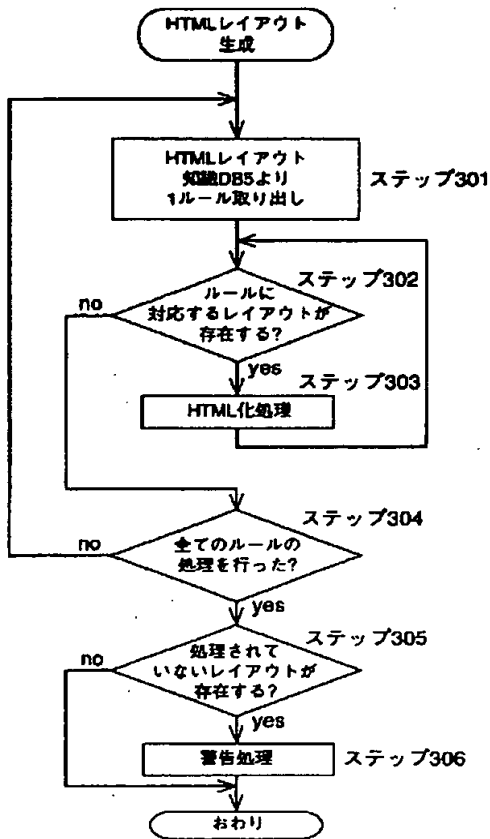


【図6】





【図4】



【図8】

我が家のレシピ集		
肉を使ったメニュー	魚を使ったメニュー	野菜を使ったメニュー
<ul style="list-style-type: none"> <li>・ 昨日のメニュー</li> <li>・ 今日のメニュー</li> <li>・ 明日のメニュー</li> </ul>		

【図7】

```

<HTML>
<BODY>

<CENTER> <H1> 我が家のレシピ集 </H1> </CENTER>

<UL>
<LI> 昨日のメニュー </LI>
<LI> 今日のメニュー </LI>
<LI> 明日のメニュー </LI>
</UL>

</BODY>
</HTML>

```

【図9】

- ・ センタリングされている大きな文字  
→ <CENTER> <H1> と </H1> </CENTER> で囲む
- ・ 箇条書きになっている部分  
→ 全体を <UL> と </UL> で囲み、  
それぞれの項目を <LI> と </LI> で囲む。
- ・ 罫線で囲まれた3段組の表  
→ 全体を <TABLE BORDER COLS=3> <TR> と  
</TR> </TABLE> で囲み、  
それぞれの項目を <TD> と </TD> で囲む。

...

【図10】

```

<HTML>
<BODY>

<CENTER> <H1> 我が家のレシピ集 </H1> </CENTER>

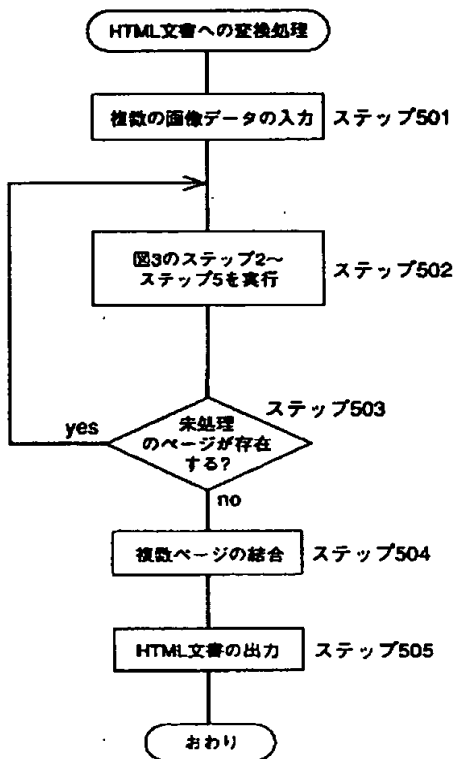
<TABLE BORDER COLS=3> <TR>
<TD> 肉を使ったメニュー </TD>
<TD> 魚を使ったメニュー </TD>
<TD> 野菜を使ったメニュー </TD>
</TR> </TABLE>

<UL>
<LI> 昨日のメニュー </LI>
<LI> 今日のメニュー </LI>
<LI> 明日のメニュー </LI>
</UL>

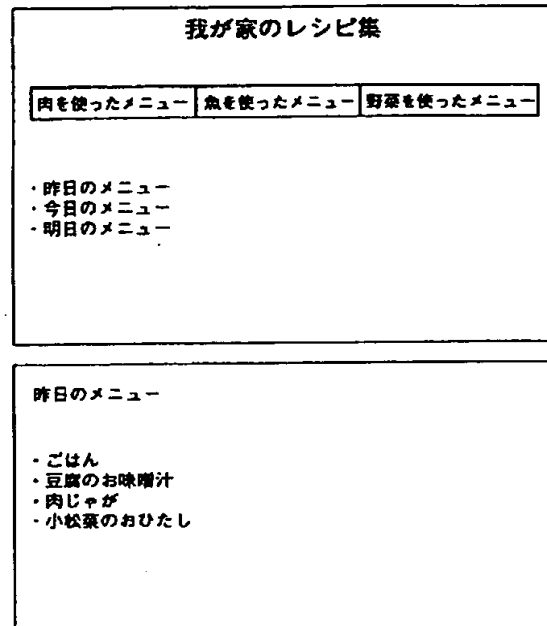
</BODY>
</HTML>

```

【図11】



【図12】



【図13】

```

<HTML>
<BODY>

<CENTER> <H1> 我が家のレシピ集 </H1> </CENTER>

<TABLE BORDER COLS=3> <TR>
<TD> 肉を使ったメニュー </TD>
<TD> 魚を使ったメニュー </TD>
<TD> 野菜を使ったメニュー </TD>
</TR> </TABLE>

<UL>
<LI> 昨日のメニュー </LI>
<LI> 今日のメニュー </LI>
<LI> 明日のメニュー </LI>
</UL>

</BODY>
</HTML>
  
```

```

<HTML>
<BODY>

  昨日のメニュー

<UL>
<LI> ごはん </LI>
<LI> 豆腐のお味噌汁 </LI>
<LI> 肉じゃが </LI>
<LI> 小松菜のおひたし </LI>
</UL>

</BODY>
</HTML>
  
```

【図14】

```
<HTML>
<BODY>

<CENTER> <H1> 我が家のレシピ集 </H1> </CENTER>

<TABLE BORDER COLS=3> <TR>
<TD> 肉を使ったメニュー </TD>
<TD> 魚を使ったメニュー </TD>
<TD> 野菜を使ったメニュー </TD>
</TR> </TABLE>

<UL>
<LI> 昨日のメニュー </LI>
<LI> 今日のメニュー </LI>
<LI> 明日のメニュー </LI>
</UL>

  昨日のメニュー

<UL>
<LI> ごはん </LI>
<LI> 豆腐のお味噌汁 </LI>
<LI> 肉じゃが </LI>
<LI> 小松菜のおひたし </LI>
</UL>

</BODY>
</HTML>
```